Establishing Pattern of Life Based on Passive Collection of WiFi Transmissions

GIAC (GAWN) Gold Certification

Author: Jason Leverton, Leverton.jd@gmail.com
Advisor: Clay Risenhoover

Accepted: January 22, 2022

Abstract

WiFi has become a staple service with the growth of home networking, smart devices, IoT, and several devices that utilize WiFi to function.

Recent attempts have been made to enhance devices due to the privacy concerns of passive WiFi collection. Location services, subject identification, or population enumeration were easily captured prior to address randomization.

However, more targeted collections methods are still available in specific applications. Certain characteristics inherently exist in the WiFi protocol that allows attribution to be made, which can be used to identify subjects, their usage on the spectrum, and even give away their physical co-location in a target area. This paper will explore the mechanisms that exist through the passive collection of WiFi signals to build a subject's Pattern of Life (PoL) that persists even through privacy enhancements of address randomization. This research will identify the attributions that remain available for passive capture and that can be used as a tool for enhanced surveillance on a subject. Conversely, the research will inform potential defensive postures to increase OPSEC and PERSEC while utilizing WiFi.

1. Introduction

Wireless connections have long been used to track presence and determine identity, volume, and attribution. The roots of wireless communication analysis go back to World War I, when the very presence or type of signals would predict or imply some type of action (Tagg, 2017). One hundred years later, the concept remains the same, analysts must try to passively collect and observe what is being exposed. Today, WiFi, or the 802.11 standard, is an incredibly widespread bearer for wireless transmissions and is still susceptible to the same challenges faced in World War I.

The 802.11 standard has had several updates to make it more privacy and security-centric. Prior to 2016, the consistent and unique address allocation in the WiFi header allowed for a straightforward mechanism for mass targeted surveillance. Individuals could be tracked across geographical boundaries by simply having the right equipment in place. These surveillance techniques also allowed population metric estimations to be made at any physical venue, whether a neighborhood, stadium or workplace. The barrier to entry for targeting was to simply listen and collect the unique 48-bit address transmitted with each wireless signal (Sapiezynski, 2015).

While it was possible to intervene and alter the unique WLAN addresses on certain platforms, the growing number of devices were not considering or were unwilling to allow the user to make these modifications. Cellular handsets, laptops, watches, vehicles, children's toys all had embedded communication capabilities, therefore it would be a daunting effort to control the wireless exposure of every device for an individual.

With the release of Google's Android 10 in 2019 and Apple's iOS 14 in 2020, the idea of address randomization was put into circulation by default. While many components are still slow to implement this new security standard, the volume of handsets targeted with these updates has removed many existing capabilities and tools and rendered them nearly obsolete for widespread collection. However, it is essential to understand the exact behaviors of these changes. Also, what aspects of the technology are so fundamental that they could not be adjusted regardless of

Jason Leverton, Leverton.jd@gmail.com

the security or privacy impact. These changes effectively tackle widespread public surveillance but do not protect against targeted surveillance. These announced and documented releases cover a specific application of address randomization. Research shows that changes were made to how a device would scan for a network (Ribeiro, 2020). Figure 1 taken from (Ribeiro, 2020) shows that as early as 2016 mobile devices were beginning to roll out address randomization, with nearly a full rollout on iOS and Android by 2019.

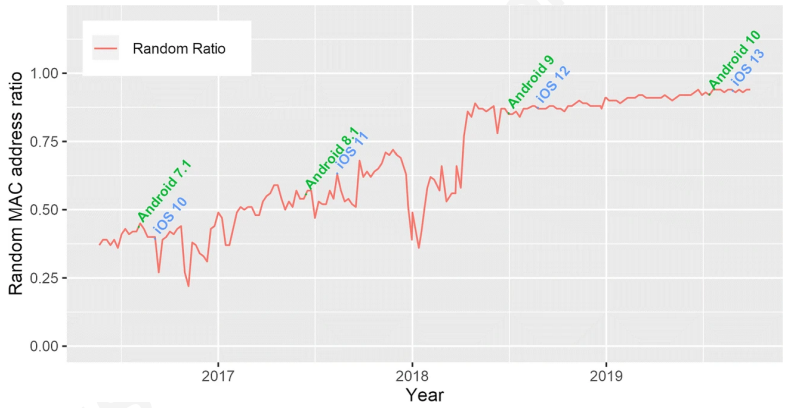


Figure 1: Ratio of Random Addresses of Android and iOS

This research will explore the 802.11 standards and construct a life pattern on a target through the passive collection of WiFi signals while considering the impact of address randomization. This paper will document the research of what is being provided to a passive collector and provide the data structures of interest. Also, thresholds and analysis will look at live traffic analysis to determine if the data is suitable to construct a PoL.

1.1 WiFi Parameters and Scope

The wireless spectrum can be abstracted to the user in two typical WiFi and Bluetooth environments. While they share similar characteristics, this paper will be specific to the IEEE 802.11 standard (IEEE, 2016).

Jason Leverton, Leverton.jd@gmail.com

The protocol and frequency are both important parts of a WLAN connection. For this paper, the 802.11 frame format is agnostic to the selection of protocol or frequency. The primary frequency and protocol used in this research will be the (2.4 GHz Channel 1) and a protocol (802.11n). The data will be consistent if applied to any other protocol or frequency that utilizes the 802.11 frame header (Shao, 2015). The one documented exclusion was the updated protocol for 802.11s for mesh networking (Wireshark, 2021).

For field applications of this approach, targeted reconnaissance will be required, and passive collectors will need to be tuned to the frequency on which the base station and target are communicating. The common WLAN configurations make use of either the 2.4Ghz or 5Ghz bands. The base station will determine the frequency and channel that will be used (Calhoun, 2009 ; Goovaerts, 2019). The output of a consolidated and truncated site survey can be viewed in Figure 2. The frequency and channel selection can be made with an active or passive scan. The results shown in Figure 2 are from a passive scan. Once the survey is complete, the target network is located, and the resulting channel is provided. If the target network is "******lie," the collection would take place on 5Ghz Ch 44. If the target network is "Bunker" then the collection would be done on 2.4Ghz Ch1.

The contents of this survey are redacted due to geo-location / SSID mapping tools such as Wigle.net.

```
SSID, BSS, RSSI, Channel, Time
[redacted]lie", "[redacted]40:1D:A4", "-90", "44", "2:15:56 AM"
[redacted]odo", "[redacted]87:2B:2B", "-83", "44", "2:14:31 AM"
"[redacted]odo", "[redacted]:A0:67:BF", "-83", "6", "2:16:04 AM"
"[redacted]Net", "[redacted]:54:D1:F0", "-96", "11", "2:15:38 AM"
"", "[redacted]A1:67:BC", "-81", "6", "2:14:37 AM"
"", "[redacted]9D:E7:30", "-91", "157", "2:15:56 AM"
"Bunker", "80:EA:96:EE:69:68", "-61", "1", "2:14:31 AM"
"[redacted]018", "[redacted]B4:23:18", "-83", "6", "2:15:56 AM"
"[redacted]Guest", "[redacted]85:38:A9", "-93", "11", "2:14:57 AM"
"", "[redacted]5F:26:43", "-88", "11", "2:15:38 AM"
"[redacted]on24", "[redacted]E9:4F:F8", "-80", "9", "2:14:31 AM"
"Sma[redacted] [redacted]E1:FC:0E", "-83", "11", "2:14:31 AM"
```

Figure 2: Base Station Site Survey Result

1.2 Pattern of Life Definition and Scope

The goal of this analysis is to establish a PoL. The following definition captures the spirit of PoL analysis.

> A method of surveillance specifically used for documenting or understanding a subject's (or many subjects') habits. This information can then be potentially used to predict future actions by the subject(s) being observed. This form of observation can, and is, generally done without the consent of the subject, with motives including but not limited to security, profit, scientific research, regular censuses, and traffic analysis. Unlike these specific areas of surveillance, pattern-of-life analysis is not limited to one medium and can encompass tracking anything in an individual's (or system of individuals') life from their internet browsing habits to their geophysical movements. (Khan (LinkedIn), 2015)

The "Pattern of Life" under this definition will be attempted on the following parameters:

1. WiFi - IEEE 802.11 Frames (Radiotap Headers)
2. All collections are done on WPA2 encrypted networks
3. All collections are passively collected with no knowledge or decryption of the frames.

1.3 WiFi Area of Effect

The analysis will be conducted on personally owned and operated devices for security and privacy reasons. The collection and retention of the data will be done according to the appropriate Canadian laws (Justice Canada, 2021). Filters and equipment were used to mitigate all possible exposures to devices not specifically included in this research. Please consult local laws when reproducing this research or conducting any of the collection methods found within this research.

For additional information and detailed threat analysis of potential issues of collection and exposure, refer to RFC 5418 (Kelly, 2009).

1.4 Address Randomization

Address Randomization will be enabled on all test Android and iOS handsets.

Android documentation states: *"MAC randomization prevents listeners from using MAC addresses to build a history of device activity, thus increasing user privacy."* (Apple, 2021)

Apple provides a better example of how the behavior of how address randomization will function:

> Because a device's MAC address changes when disconnected from a WiFi network, it cannot be used to persistently track a device by passive observers of WiFi traffic, even when the device is connected to a cellular network. (Google, 2021)

These specific updates bring the control of address randomization to the user. Before these announcements, both Apple and Google were rolling out address randomization functionality to the probe request/response frames used in network discovery.

Past privacy changes were transparent to the user and were seen in the wild as early as 2016 (iOS/Android) (Ribeiro, 2020). The recent adjustments to address randomization are exclusive to the behaviors when the device connects to a WiFi network for the first time. Each network will utilize a new randomized address. This new address becomes fixed once established, and will remain fixed while the device has that connection profile saved. Fixed random addresses are the most important aspect of address randomization functions.

1.5 Equipment and Software

The following contains the components that were used in this research. With the differences in hardware and software throughout the industry, it will be noted when there is an

expectation that there will be a specific platform or version dependency on any indicators or data sets.

Collection Antennae

- Mac OS X Macbook Pro
- Alfa AC1900 WiFi Adapter

WiFi Base Station:

- Apple Airport Extreme AC

Cellular Handsets

- Apple iPhone 11 Pro running iOS 15
- Samsung Galaxy S9 running Android 10

Collection Software

- TShark
- Splunk

2. IEEE 802.11 Utilization

The IEEE 802.11 standard will be a key component in this research. This section will cover the protocols, sub-standards and key materials that will be used.

2.1 Frame Addresses

48-bit addresses will be used in the following format (Shao, 2015):

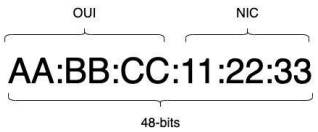


Figure 3: Address Format

The organizationally unique identifier (OUI) portion of the address refers to the manufacturer of the transmitter, and the OUI values are registered and can be queried with open databases. Figure 4 shows that the transmitter OUI from a Macbook Pro is registered to Apple.

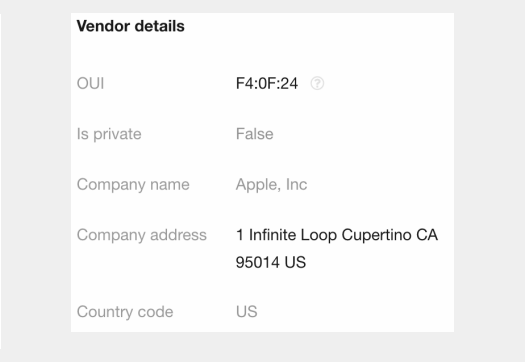


Figure 4: Address with registered OUI

However, with address randomization, the behavior varied between Apple and Samsung handsets. A bulk lookup tool showed that no address in use on an Apple device with randomization was registered (Wireshark, 2021). The general observation is that the address randomization is defined to exclude already registered OUIs. Referencing open registration lists such as the one at http://standards-oui.ieee.org/oui28/mam.txt, duplications of a random address and a registered OUI space for iOS could not be found. Samsung random addresses assigned when joining networks were still registered with Samsung.

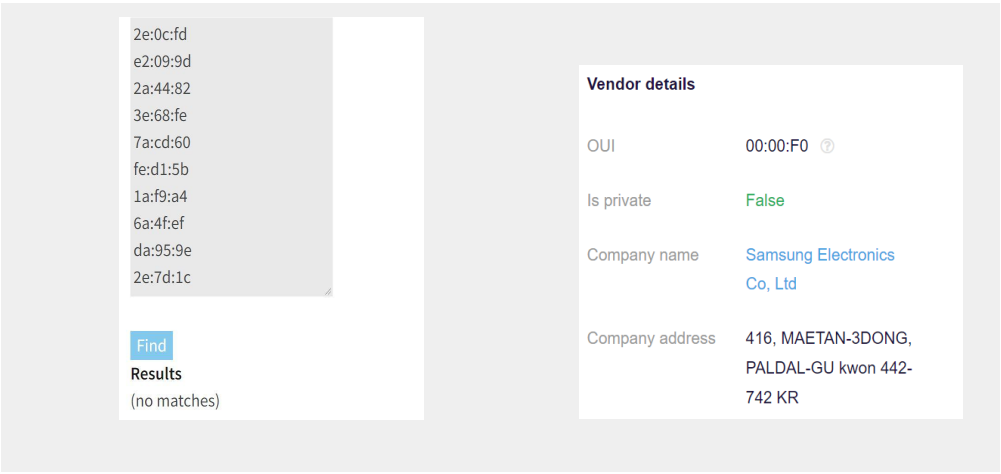


Figure 5: Randomized address OUI Lookups (Apple Left, Samsung Right)

2.2 Frame Types

IEEE 802.11 frame types come in three planes (Calhoun, 2009):

1. Management
2. Control
3. Data

There are over 40 types, and not all apply to this context. The header structure is discussed in detail at RFC5416 and enumerated well at the cited reference (Darchis, 2020).

Throughout the paper, the following types will be of interest, but (Darchis, 2020) should be consulted for a full list of types:

- Management - Disassociation (0x0a): This will provide context on the parameters around the device exit from the network.
- Management - Authentication (0x0b): This will provide the authentication challenge when a device joins a network.
- Management - Deauthentication (0x0c): Not expected to show up as a meaningful metric, but interesting to keep an eye on as it could represent environmental challenges or a hostile wireless spectrum
- Data - Data (0x20) - Frame containing data
- Data - QoS Data (0x28,0x29,0x2a,0x2b) - May be present depending on device / base station

The data structures are slightly offset from their enumerated values in (Darchis, 2020) as they show up in the stream. The data frame 0x20 would correspond to the second most significant digit. The flags set can provide some unique indication of behavior. Figure 6 shows that the data type is part of the Frame Control Field and is represented in this context as the digit "8".

Jason Leverton, Leverton.jd@gmail.com

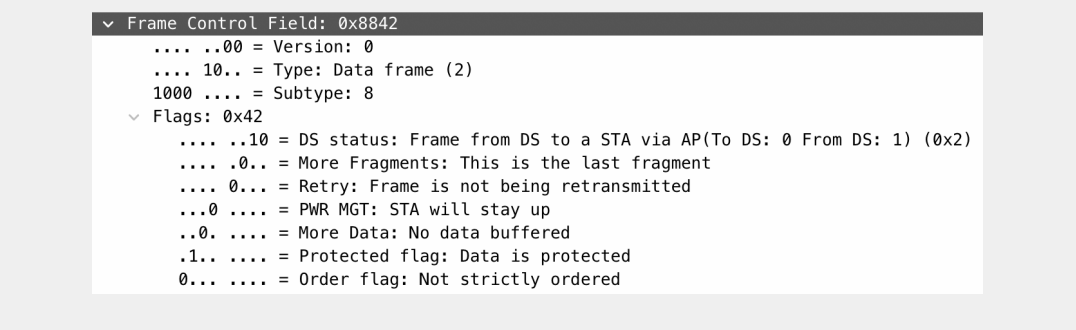


Figure 6: Data structure of 802.11 Types in Capture

2.3 Protocols

The protocols will not alter the analysis. However, the protocols can potentially dictate how the radiotap headers are collected. In some protocols, the bandwidth will be adjusted, and the capture antenna must be compatible with the base station configuration to collect the radiotap headers passively (Alfa, 2021).

The collection antennae utilized in this paper are compatible with protocols A, B, G, N, and AC. Others may also be appropriate, but they were not considered or analyzed for compatibility.

2.4 Spectrum

The 5Ghz spectrum has a large variety of applications depending on the country, protocol, and usage of the transmitter (Wikipedia, 2021). With the 802.11n protocol, the channels in use are:

- 2.4 GHz; Channel 1
- 5 GHz; Channel 64

2.5 Data Flow

The data flow of the passive capture is shown in Figure 7. The collection was conducted on radiotap headers and converted to .pcapng files. The collection built the .pcapng file for 15 minutes, then committed the file. Once the file is saved to disk, tshark is used to extract the data

out of the file into a more digestible tab-separated value (tsv) format. The selected fields and their output formats are shown in the table below. Converting the data to a structured format is necessary as it transforms a dataset of Gigabytes to Kilobytes. The approximate transform ratio is 200:1. Additionally, capture filters were placed on the radiotap to isolate traffic destined to the base station or the target.

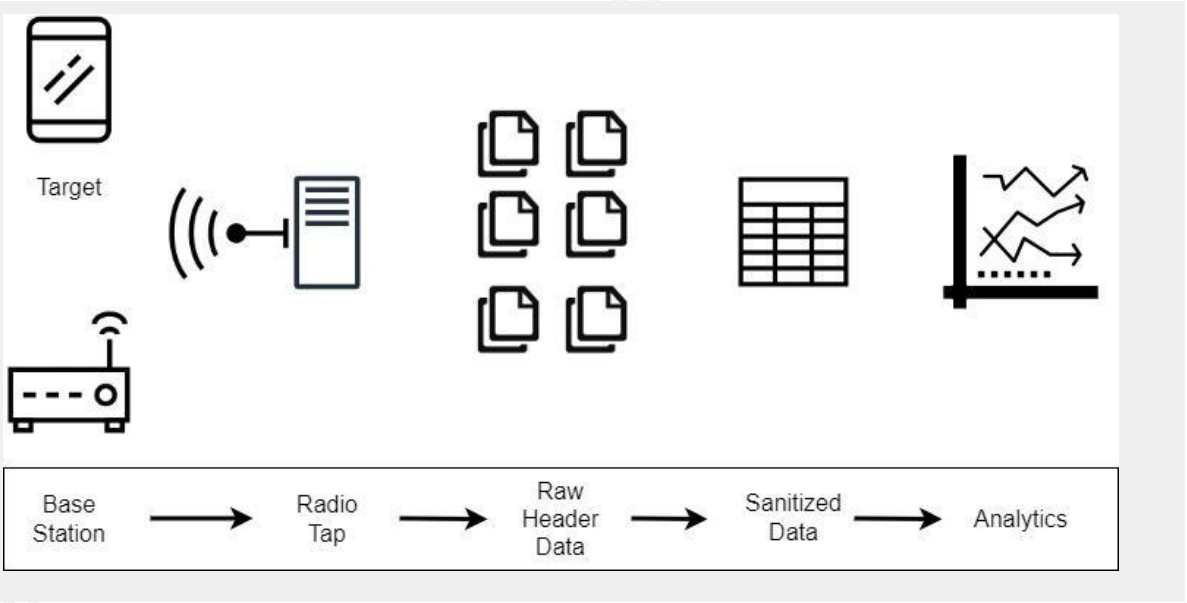


Figure 7: Collection Data Flow

The structure of the .tsv file is shown in Table 1, with the corresponding tshark identifiers. These values were selected based on the value of the information they contained. In some cases, they evolved based on data challenges or erroneous formatting. One example of this was the filter for the subtype. Wireshark/Tshark fields will identify the type through their lookup. However, when the filter WLAN.fc.type_subtype was used, all the type values were exported as 0x28. The variety was lost in export. The only way to overcome this was to export the entire 4-byte WLAN.fc filter and perform the lookups manually (Wireshark, 2021). A second example was the inclusion of noise in the data set. There were very strange patterns in the signal strength on some captures that corresponded to a large increase in noise. A signal-to-noise ratio will make

the results much more consistent. Epoch time was used since this was the easiest format for Splunk to translate the time on import.

_time	wlan.src	wlan.dst	noise	signal	length	type
1638131525.63	2e:0c:fd:c1:59:ed	80:ea:96:ee:69:68	-101	-44	53	0x0008801

Table 1: Sanitized PCAP Data Structure

The command that prepares this data is (Wireshark, 2021):

```
Tshark -r <capture_file> -T fields
        -e frame.time_epoch
        -e wlan.ta -e wlan.da
        -e radiotap.dbm_antnoise
        -e wlan_radio.signal_dbm
        -e frame.len
        -e wlan.fc.type_subtype
```

-T : sets the format for the output
-e : selects the fields that will be extracted

2.6 Passive Collection

With the nature of the passive collection, the environment may not always produce consistent results with the target and base station or even with the target over time. The passive collector can only capture a frame that can separate the signal from the noise. Three conditions may exist when isolating the target and its pair base station.

- Condition 1 (C1): Collector can collect Target and Base Station
- Condition 2 (C2): Collector is only able to collect from the Target
- Condition 3 (C3): Collector is only able to collect from the Base Station

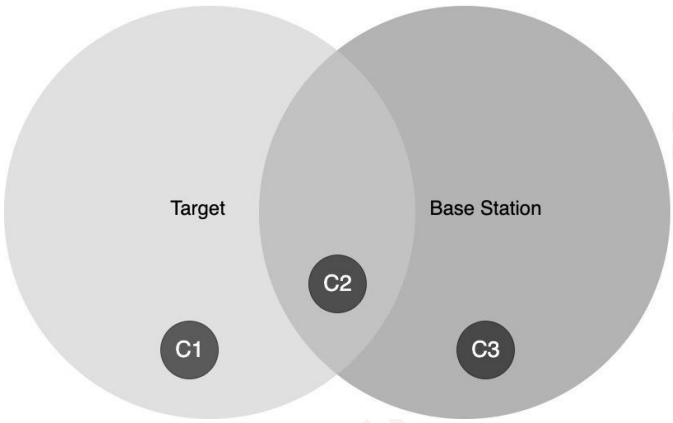


Figure 8: Wireless Area of Collection Conditions

The three conditions are each able to provide some element of behavior. C2 is the ideal condition as it allows the capture of both the target and base station, giving a much more accurate picture of geo-location and data usage. C1 and C3 can still provide a window of behavior but will require that analysis be adjusted on whether the frame contains receiving or transmitting sources.

2.7 Address Randomization Behavior Analysis

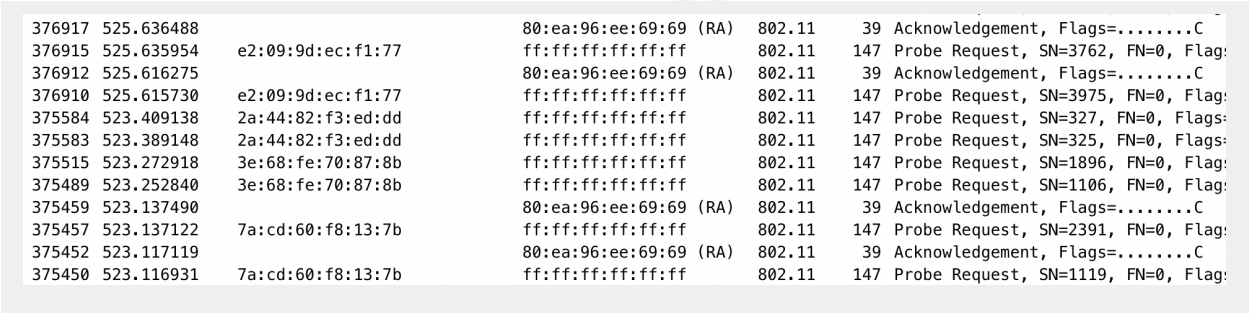
Before indicators and analyses on patterns of life can be extracted, it is necessary to understand exactly how the randomization occurs. A few methods can be used to observe the connection behaviors with address randomization enabled.

There are three observable states that a device exhibits:

1. Not connected to a WLAN and actively searching
2. Connected to a WLAN
3. Off

The state of actively searching when the device actively searches for a network to join. This scan is how the device collects the list of WiFi access points available to be joined. The

device and base station communications are primarily probed request/response subtypes. These messages are constantly being transmitted from the device in this state. Figure 9 shows a sample capture from a single mobile device during this active scan phase. The source transmissions in column 3 show that each probe request adjusts the transmission address after two transmissions. This randomization behavior makes it near impossible to pinpoint a source transmission when a device is in this state. Controlling the signal isolation was the only way to isolate this traffic to ensure it was all originating from the same device.



376917	525.636488		80:ea:96:ee:69:69 (RA)	802.11	39	Acknowledgement, Flags=........C
376915	525.635954	e2:09:9d:ec:f1:77	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=3762, FN=0, Flags
376912	525.616275		80:ea:96:ee:69:69 (RA)	802.11	39	Acknowledgement, Flags=........C
376910	525.615730	e2:09:9d:ec:f1:77	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=3975, FN=0, Flags
375584	523.409138	2a:44:82:f3:ed:dd	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=327, FN=0, Flags=
375583	523.389148	2a:44:82:f3:ed:dd	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=325, FN=0, Flags=
375515	523.272918	3e:68:fe:70:87:8b	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=1896, FN=0, Flags
375489	523.252840	3e:68:fe:70:87:8b	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=1106, FN=0, Flags
375459	523.137490		80:ea:96:ee:69:69 (RA)	802.11	39	Acknowledgement, Flags=........C
375457	523.137122	7a:cd:60:f8:13:7b	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=2391, FN=0, Flags
375452	523.117119		80:ea:96:ee:69:69 (RA)	802.11	39	Acknowledgement, Flags=........C
375450	523.116931	7a:cd:60:f8:13:7b	ff:ff:ff:ff:ff:ff	802.11	147	Probe Request, SN=1119, FN=0, Flags

Figure 9: Randomization in Probe Request/Response Frames

The second stage is when a device connects to a wireless network. The behaviors in this state are much more interesting and must respect network constraints. Frequent randomization on transmission addresses would put significant pressure on arp tables, wireless controllers, and DHCP servers, so it is understandable that device manufacturers approached this stage with compatibility as the priority and eased off the requirements for privacy.

This functionality informs one of the key behaviors that allows all other behaviors to be derived. Once a device is authenticated on a network, it will randomize the address at the initial network-join and then maintain that address for that SSID pairing indefinitely until the network is forgotten. This exchange offers two opportunities to isolate a target device on a network.

If the target can be observed entering the network boundary, capturing all Extensible Authentication Protocol over LAN (EAPoL) frames can strongly correlate to network active and physical presence. EAPoL packets are only transmitted during the initial joining and

authentication on the network. EAPoL functionality makes the presence of these packets rare in comparison. The rarity of EAPoL packets can then assist in attributing the arrival of a target.

If an EAPoL packet is missed or not correctly attributed, the second method is to monitor the base station for all addresses to which the base station communicates. The base station will only communicate with connected peers, and if a device is not on the network, the address will not be available. Once a device joins a network with the base station, it will begin communicating. So only two data sets are required. The first data set will contain a collection of traffic when the target is not in the area of collection. The second data set will contain a collection of traffic when the target is in the area of collection.

```
150230 212.723068    98:52:4a:b5:77:80          a2:a9:31:84:8a:c2       EAPOL     162 Key (Message 1 of 4)
150232 212.726898    98:52:4a:b5:77:80          a2:a9:31:84:8a:c2       EAPOL     218 Key (Message 3 of 4)
```

Figure 10: EAPoL Frames

3. Pattern of Life Parameters

3.1 Target isolation

Target isolation with address randomization does require a second source of information to make the necessary link to acquire the target address. Two examples are suggested below, but other options can exist, and even techniques discussed further in this paper can be used to fingerprint users on networks without knowing the target address.

1. EAPoL packets are linked to the target entering the passive collection area. E.g
    a. Target is observed entering an area at 15:03 EST.
    b. A single EAPoL is observed between 15:00 and 15:30 EST.
    c. Attribution can now be assigned based on the pairing of these two events.
2. Base station communications are isolated to include two sets—one set where the target is not communicating with the base station, and another instance when they are. The difference in the two lists provides a small subset to attribute to the target.

Once attribution is made, analysis and behavior can increase future attribution through other behavior mechanisms such as active use time, signal strength, and connection with other devices. The filters for these options are:

Display filter for EAPoL packets

```
eapol
```

Capture filter isolating the base-station

```
wlan.ta <base station address> or wlan.ra <base-station address>
```

3.2 Target Enters an Area

A target entering an area is a well-defined and specific action and difficult to obscure.

EAPoL Frames and target_address filtering can signal the moment the target has entered the area with a high degree of reliability. Once the target has been identified, additional filtering can lower collection rates when the target is not in the area.

Target communications:

```
wlan.ta == <target_address> || wlan.ra == <target_address>
```

EAPoL frame:

```
eapol
```

3.3 Target Departs an Area

There are two ways a target can exit a network (Shao, 2015):

1. Gracious - Target will send a Disassociate message to the base station informing that it will no longer use the connection. The disassociate type frame is paired with the following actions:
    a. WiFi antenna is turned off in the OS, or Airplane mode is activated.
    b. The target joins a different WiFi network.
    c. The phone is powered off normally.
2. Ungracious - Target leaves the area without announcement and without notice.
    a. Target is no longer in range of the base station.
    b. System failure. (power cycle, crash, reset)

Disassociate frames do not provide a window to the physical state of the target. These frames would be good to collect and analyze if the target is applying mitigating behaviors or utilizing other base station connection points. An interesting observation in this condition is that device attribution can be made by capturing connection disassociation. Samsung handsets apply a tag of data specifically on this dissociated frame. Android and iOS messages are shown below, respectively.

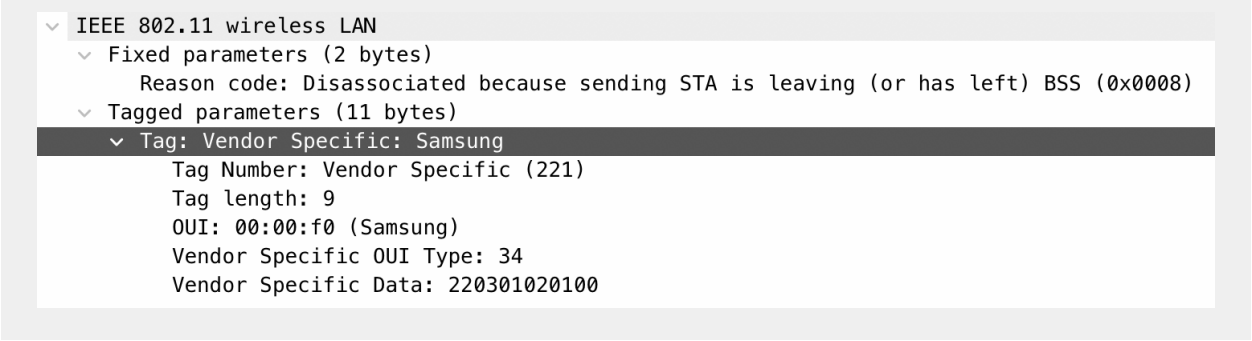


Figure 11: Samsung Disassociation Additional Tag

Figure 12: Apple iOS Disassociation

The best indication of the target departing the area is the increase in signal on frames collected from the target, the absence of frames collected from the target address, and the series of retransmissions from the base station. These factors combined can establish an accurate threshold for the target departing an area. The filter for the target should only be based on the transmitted frames from the target. The base station, bonjour services, and other devices make the inclusion of the receiving address of the target an unreliable metric.

The display filters can be set as:

Disassociation frames:

```
wlan.fixed.reason_code == 0x0008
```

Target Isolation:

```
wlan.ta == <target_address>
```

Signal Isolation:

```
radiotap.dbm_antsignal > -30 (-20 to -80)
```

3.4 Target is in a State of High/Low Usage

The availability of the transmissions on a WiFi network can provide a window into the usage patterns of the target. The frames provide frequency, length, and even type for each frame

transmitted and received by the target. Aggregated frames will be used to build profiles and thresholds for observed conditions.

The data set for this collection will be very large, even with isolated captures. In order to create a more comparable dataset, the set will be grouped into time buckets, which will establish a more resilient trendline and lower the false positives of thresholds. The resolution of the bucket can vary between one second and one hour. A time window of 5-15 minutes removed much of the noise from the data but still highlighted important events and trends.

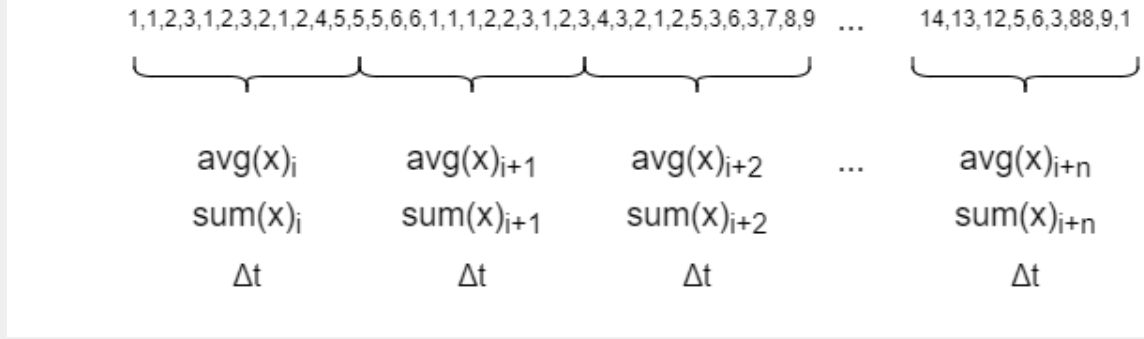


Figure 13: Data Set Management Overview

The potential use cases from the information available within the frames could be:

- Over time, the plot of the signal can provide information on the target's physical location.
- The plot of data frames (frame types described in Section 2.2) over time can provide an indication of the volume of data flowing into and from the target.
- The volume of data frames indicates the flow of information, whether uploads or downloads are being performed, and the direction of the flow of that information.

Base station collection can be done to capture/obtain a sense of variability on the signal with the collector. If the collector is collecting from a fixed location, this can provide context on

the relative strength of each packet. This information could then be used as an accurate threshold for other targets. A base station typically is a fixed appliance in the home. When analyzing the received signal from the base station, a threshold can be derived from the standard deviation of the signal. The signal remains at a relatively constant db value with a low standard deviation. In testing, it was noted that 2.93 or lower was a good threshold for the standard deviation of a static transmitter. Even with two fixed sources, the signal is variable due to physical and electronic interference, so it is useful to know how variable that signal is within the environment.

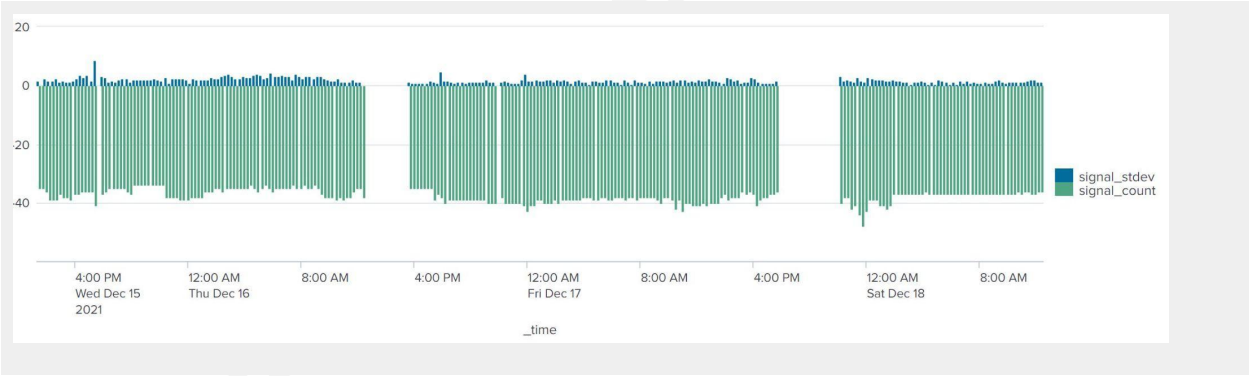


Figure 14: Base Station Signal Timechart

```
index="index_d" sourcetype="tsv" wlan_ta="80:ea:96:ee:69:68"
| bucket _time span=15m
| stats stdev(signal) as signal_stdev avg(signal) as signal_count by
_time
| eval signal_count=round(signal_count)
| eval signal_stdev=round(signal_stdev,1)
```

4. Pattern of Life Analysis

The analysis of the PoL will follow a collection of approximately 72 hours and will be aligned to the real-world events as they were recorded. The geographical and usage events captured outside the passive collections are listed in Appendix A.

4.1 Attribution Indicators

4.1.1 Target Enters an Area

Table 2 shows the collection of all EAPoL connections taken during the collection period.

Event	System Time	Length	Type
1	Dec 17 2021 10:11:37 PM	162	0x00008802
2	Dec 17 2021 10:10:09 PM	162	0x00008802
3	Dec 17 2021 10:10:08 PM	162	0x00008802
4	Dec 17 2021 10:10:07 PM	162	0x00008802
5	Dec 17 2021 10:10:06 PM	162	0x00008802
6	Dec 17 2021 3:43:46 PM	162	0x00008802
7	Dec 17 2021 1:56:20 PM	162	0x00008802
8	Dec 16 2021 10:12:24 PM	162	0x00008802
9	Dec 16 2021 3:42:37 PM	162	0x00008802

Table 2: Collection of EAPoL Packets

For the sake of example, the physical target was confirmed to arrive in the area at 15:32 on December 16th (event #9). The association to the target can be set to address 2e:0c:fd:c1:59:ed.

```
Frame 3: 218 bytes on wire (1744 bits), 218 bytes captured (1744 bits) on interface 0
  Interface id: 0 (en0)
  Encapsulation type: IEEE 802.11 plus radiotap radio header (23)
  Arrival Time: Dec 16, 2021 15:42:37.957642000 EST
  [Time shift for this packet: 0.000000000 seconds]
  Epoch Time: 1639687357.957642000 seconds
  [Time delta from previous captured frame: 0.012677000 seconds]
  [Time delta from previous displayed frame: 0.012677000 seconds]
  [Time since reference or first frame: 0.014903000 seconds]
  Frame Number: 3
  Frame Length: 218 bytes (1744 bits)
  Capture Length: 218 bytes (1744 bits)
  [Frame is marked: False]
  [Frame is ignored: False]
  [Protocols in frame: radiotap:wlan_radio:wlan:llc:eapol]
Radiotap Header v0, Length 25
802.11 radio information
IEEE 802.11 QoS Data, Flags: ......F.C
  Type/Subtype: QoS Data (0x0028)
  Frame Control Field: 0x8802
  .000 0001 0011 1010 = Duration: 314 microseconds
  Receiver address: 2e:0c:fd:c1:59:ed
  Transmitter address: 80:ea:96:ee:69:68
  Destination address: 2e:0c:fd:c1:59:ed
```

Figure 15: EAPoL to Target Attribution Frame

4.1.2 Target Departs an Area

The clearest indicator of the target no longer in the area is the absence of frames picked up by the collector originating from the target. It is easy to spot the large departures from the area on a time chart, but difficult to determine the exact time frame accurately and identify smaller windows of departure. Figure 16 shows that three events can be identified visually but misses four smaller windows.

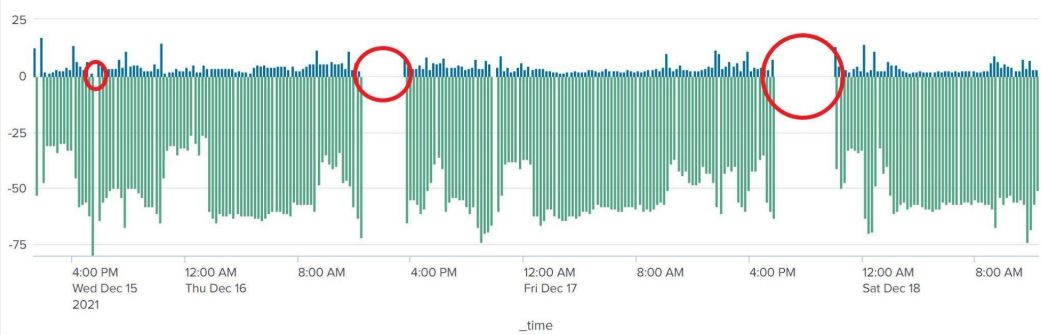


Figure 16: Target Out of Area Timechart Analysis

The event identification is better applied as a threshold within Splunk, where the time chart populates no records for a given search parameter.

```
index="index_d" sourcetype="tsv" wlan_ta="2e:0c:fd:c1:59:ed"
| timechart useother=f usenull=f span=15m(5m) count
| where count = 0
```

The time is listed as the more accurate window defined by the 5m result. The 15-minute resolution was unable to catch shorter departures properly and was not as precise as to locate the departure time from the area. For the analysis, context can be added by including multiple resolutions.

Event	Time	5 Minutes	15 Minutes	Description
1	Dec 15 @ 1330	Yes @ 1330	Yes @ 1330	Out of Area
2	Dec 15 @ 1720	Yes @ 1720	Yes @ 1730	Out of Area
3	Dec 15 @ 2330	Yes @ 2330	No	Phone Update
4	Dec 16 @ 1210	Yes @ 1210	Yes @ 1215	Out of Area
5	Dec 16 @ 2140	Yes @ 2140	Yes @ 1240	Out of Area
6	Dec 17 @ 1340	Yes @ 1340	No	Out of Area
7	Dec 17 @1535	Yes @ 1535	No	Out of Area
8	Dec 17 @ 1740	Yes @ 1740	Yes @ 1745	Out of Area

Jason Leverton, Leverton.jd@gmail.com

Table 3: Time Resolution on Captured Events

As mentioned in the 1.6 Passive Collection Setup, the target may not be in the range of the collector but could still be in the range of the base station. If this situation occurs, communications from the base station could produce meaningful data. When analyzing the frame type, there is a unique pattern when the base station attempts to retransmit frames and the incoming new data frames approach 0. The unique presence of event 0x884a (802.11 Data Frame with the retransmission flag set) being at the top of the types observed in a given time bucket is unique to a target no longer in range of the base station. The secondary bars are type 0x0008842.

Figure 17: Target Departing Area Threshold

4.2 Behavioral / Statistical Indicators of Interest

4.2.1 High/Low Usage

The volume of connections is another difficult metric to hide. The data contained in the frames are likely secured with multiple layers of encryption. However, each byte of that data is on the capture. For the purpose of determining indicators of interest, the data sets that will produce the most interesting results will be length and type. The length is the reported length of the frame. If the sum of frame lengths is taken over a specified time, the bandwidth of that connection can be determined. If the bandwidth can be accurately plotted, an indicator of usage can be derived.

The measurement of 802.11 frame types will provide a baseline that can be expected on the target device. Figure 17 contains a three-hour window where the target device had no contact at all but remained idle. The main observation with this chart is that the control frames are the prominent frames detected in this low-use window.. Data spikes like the one shown at 5:30 AM seem to be infrequent during these periods but could depend on specific device settings such as background app refresh and push settings on message retrieval services. This connection pattern would be a deliberate baseline activity that would need to be done with each device during a period where the device's state is assumed to be idle.

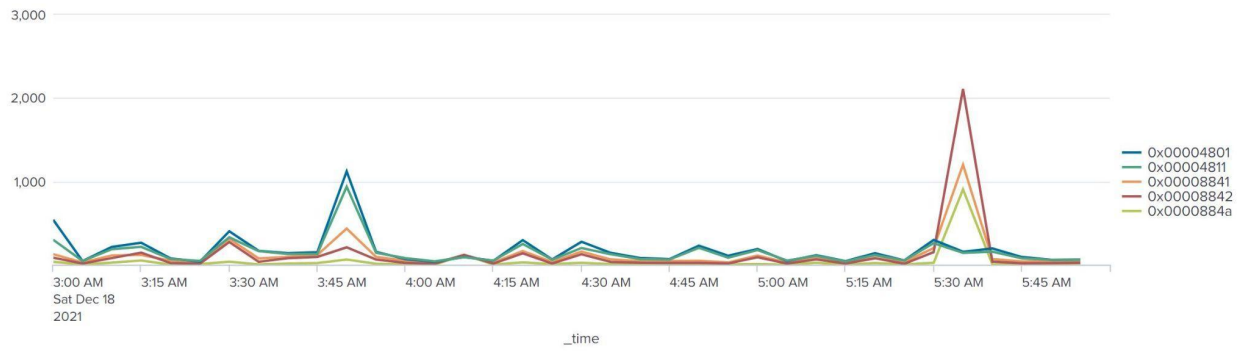


Figure 17: Timechart of Type During Low Usage

In terms of the volume of data, there are many ways to analyze the data. The collection can plot the upload and download data separately or aggregate. Figure 18 shows the time chart of the volume of data going to and from the target device throughout the collection. The data peaks were truncated to keep the scale of the time chart meaningful. The purpose is not to determine the peaks but rather to achieve a threshold of when the phone is active. The key observation on this data is that the phone receives more data than it sends with general use. Therefore, thresholds that monitor target devices could also establish thresholds on the upload and download thresholds of the device as well as their ratios. Another observation is that when the device is potentially not in use, the data throughput is much lower. The ratio of these values offers the potential for a very meaningful threshold that can assist in building a PoL metric that

can be associated to a specific activity. E.g. Once a day data upload to the cloud indicating a backup or batch file transfer.

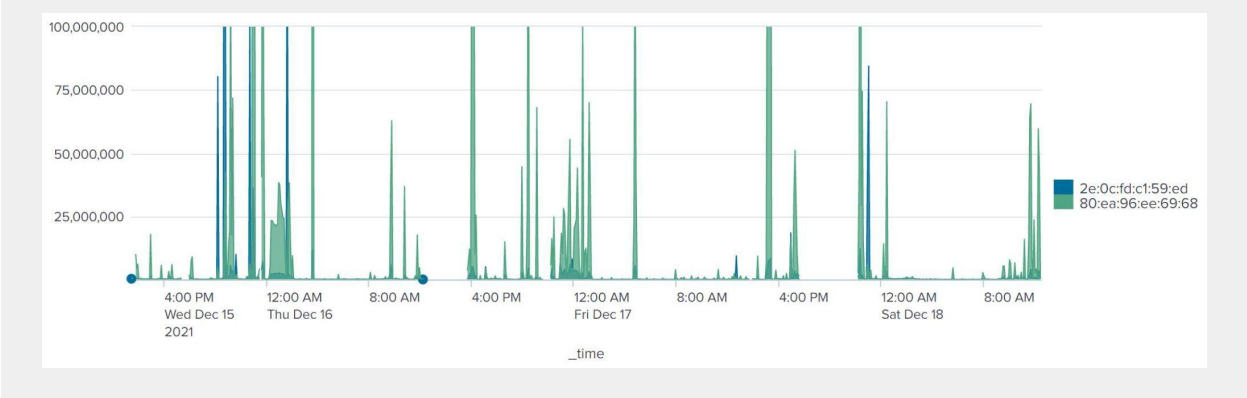


Figure 18: Upload and Download Volume Timechart

Established baselines with three different types of activities for 15 minutes were completed to set a baseline. Repeated the baseline several times and calculated an average with a metric of Kbps as a result.

The three baselines were:

1. Idle: Phone is not touched or activated in any way for 15 minutes
2. Light use: Social media, news, text-based, asynchronous content for 15 minutes.
3. Streaming: Utilizing streaming services and watching content with video/audio for 15 minutes.

The results were:

- Idle = 3.17 Kbps
- Light use = 4215.32 kbps
- Streaming = 10078.85 kbps

These results can help understand the type of usage pattern that is being observed. The one interesting observation that came up in this analysis was the unique nature of the traffic as it was observed. The consistent and frequent traffic spikes of streaming services or the way a streaming service buffering algorithm can all be detected in the time chart. Encrypted traffic analysis is a topic on its own, and WLAN encrypted traffic certainly overlaps with already existing work done on encrypted data on VPNs (Lashkari, 2016).

Three sample time charts of each occurrence are included in Appendix B.

4.2.2 Signal Strength

The value of the signal strength provides a relative measurement of the distance the signal must travel from the source. This does not always indicate a direct correlation to physical distance, so some interpretation is required.

The plot of signal strength is paired with the standard deviation of those measurements within the block of time. The purpose of this is to assess the variability of the signal. The observation is that the lower the standard deviation, the more static the device is during that collection window. Figure 14 in Section 3.4 can also be used in this application to provide the expected standard deviation that can be expected from a fixed target. The results from Figure 14 show that a standard deviation of 2.93 was calculated. When applied to data presented in Figure 19, we can see that there are three distinct locations where the signal consistently produced values with a standard deviation of three or lower.

- Location A: 0100-0800 - A signal of -61 dB
- Location B: 1000-1600 - A signal of -25 dB
- Location C: 0800-2200 - A signal of -44 dB

The results demonstrate that there are distinct location patterns in the signal reports. However, the “Pattern of Life” analysis will be dependent on the culture and lifestyle of the target. In this specific case, the observed locations correspond to Location A is sleeping.

Location B is office/work time. Location C is the common area (Kitchen, Living Room, Bathrooms). High standard deviations were most captured during mealtimes or with arrivals and departures from outside the home.

A good reference for signal strength and relative distance (Sonicwall, 2019):

- Signal < -90 dBm: this signal is extremely weak, at the edge of what a receiver can receive.
- Signal > -67dBm: this is a fairly strong signal.
- Signal > -55dBm: this is a very strong signal.

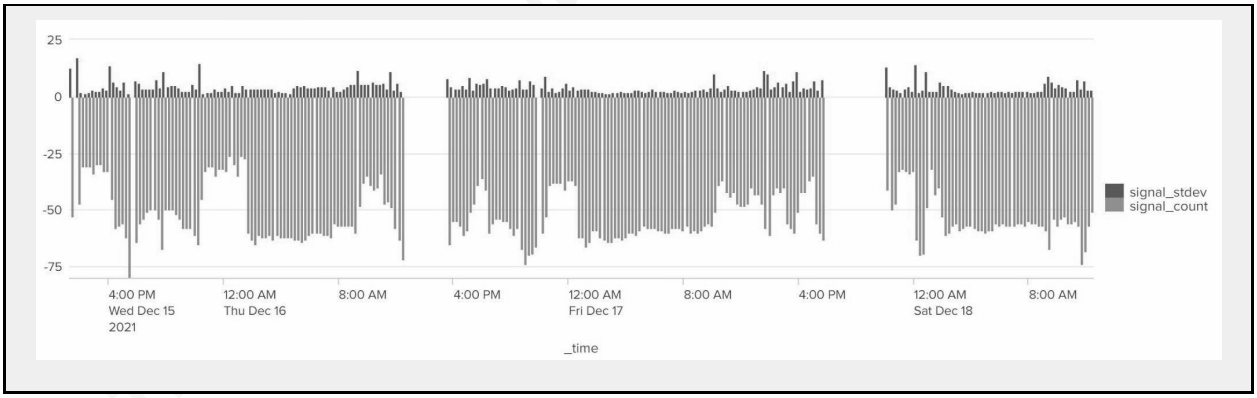


Figure 19: Timechart of Received Target Signal Strength

4.2.3 Data Noise & Other Signals

Other mentions in the data were not significant on their own but did indicate potential behaviors on the network or exclusions that were eventually filtered out completely.

1. Apple Handoff was especially verbose with multiple devices. MacBook, iPad, and iOS watches all generated traffic when associated with a target. Google hardware and services were also present. While not a specific indicator in behavior, this was an observed area of interest for future study.

2. Multicast addresses were very present on initial captures. Bonjour and other device discovery applications are very verbose and uniquely identifiable, even on an encrypted network. RFC 7042 can be referenced to help pinpoint address purposes (Eastlake, 2009).
3. Networks have bit errors, and when dealing with 10 billion packets, an event that has a 1 in a billion chance of occurring may occur ten times. The passive filter collects frames that have bit errors and can adjust the transmitter or receiver addresses by a single bit.
    a. Sample error addresses captured:
        - 80:aa:96:ee:69:68
        - 80:6a:96:ee:69:68
        - 88:ea:96:ee:69:68
        - a0:ea:96:ee:69:68

5. Conclusion

The overall results of this research demonstrated considerable gaps that still exist with passive wireless collection on handsets. Address randomization increases the privacy threshold in mass tracking efforts. However, the need to be compatible with most home network configurations means that address randomization in targeted and active usage scenarios does little to protect the individual. This research did not expose a new threat introduced by address randomization, but with the ease of collection in the past, the current bar for those interested in passive scanning still presents some options for collection. In this research, the objective was to determine if there was passive exposure and to define what is and is not in bounds with the recent changes to address randomization.

The following questions lead the research at the outset:

Objective: Was the collection able to determine if the subject was detected entering/leaving the area?

The events that showed the target entering and leaving the area were identifiable. The entering of an area is an undeniable EAPoL exchange. This behavior is the default in both Apple and Android operating systems. However, the presence of authentication cannot be associated with the target's arrival. The authentication is triggered for various reasons, including but not limited to: system reboot or joining a different network.

The target departing an area was also detectable. The absence of the target transmitter is the most unambiguous indication that the target is no longer in the area. Dissociation packets were generally associated with manipulating the handset for various reasons, including but not limited to: system reboot, joining a different network, airplane mode, or deactivation of the WiFi transmitter.

Objective: Was the collection able to determine if the subject's usage patterns were active/inactive on the network while in location?

The signal, frame size, and subtype all provided a clear view of the activity of the network. If the target was in the area and connected to the network, it was possible to make inferred conclusions on the device's state. 15-minute intervals would vary considerably on active and idle use, e.g., low use (401KB) and high use (1.36GB). The subtype allowed for the determination of the type of traffic. 802.11 frames are tagged by management, control, and data frames. The volume of frames could provide an estimate but isolating the frame type allowed a more precise determination to be made.

Objective: Was the collection able to make any inferences on PoL such as bedtime, mealtime, or consistent high/low usage times?

Periods of usage, a change in signal position and strength, and time of day all demonstrated that something was likely to have happened in that period of time. These results are highly influenced by culture and lifestyle. Shift workers, night workers,

different devices in the home, or even how the home is structured can influence the inferred data. The more data collected over days and weeks reinforces the data over time. If the target was observed to have been active throughout the afternoon and evening and then suddenly goes quiet around midnight, it strongly indicates a bedtime capture. If moderate to light consistent device usage over several days is available, definite patterns can be identified. In a three-day capture, the bedtime, wakeup time, expected departures, mealtimes, and locations in which they all take place could be derived.

References

Tagg. (2017). *Tactical signals intelligence originates in World War I.* Retrieved January 22, 2022, from https://www.army.mil/article/191282/tactical_signals_intelligence_originates_in_world_war_i

Sapiezynski. (2015). *Tracking Human Mobility Using WiFi Signals.*
Retrieved January 22, 2022, from
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0130824

Ribeiro. (2020). *Passive Wi-Fi monitoring in the wild: a long-term study across multiple location typologies.*
Retrieved January 22, 2022, from
https://link.springer.com/article/10.1007/s00779-020-01441-z

IEEE. (2016). *IEEE 802.11-2016.*
Retrieved January 22, 2022, from https://standards.ieee.org/standard/802_11-2016.html

Shao. (2015). *RFC 7494 - IEEE 802.11 Medium Access Control (MAC) Profile for Control and Provisioning of Wireless Access Points (CAPWAP).*
Retrieved January 22, 2022, from https://datatracker.ietf.org/doc/html/rfc7494

Wireshark. (2021).*File: mesh.pcap.*
Retrieved January 22, 2022, from https://wiki.wireshark.org/SampleCaptures

Calhoun. (2009). *RFC 5416 - Control and Provisioning of Wireless Access Points (CAPWAP) Protocol Binding for IEEE 802.11.*
Retrieved January 22, 2022, from https://datatracker.ietf.org/doc/html/rfc5416

Goovaerts. (2019). *Improving Privacy through Fast Passive Wi-Fi Scanning.*
Retrieved January 22, 2022, from https://papers.mathyvanhoef.com/nordsec2019.pdf

Khan (LinkedIn). (2015). *Cyber Pattern-of-Life Analysis.*
Retrieved January 22, 2022, from https://www.linkedin.com/pulse/cyber-pattern-of-life-analysis-ely-kahn

Justice Canada. (2021). *Interception.*
Retrieved January 22, 2022, from https://laws-lois.justice.gc.ca/eng/acts/c-46/section-184.html

Kelly. (2009). *RFC 5418 - Control And Provisioning of Wireless Access Points (CAPWAP) Threat Analysis for IEEE 802.11 Deployments.*
Retrieved January 22, 2022, from https://datatracker.ietf.org/doc/html/rfc5418

Apple. (2021).*Wi-Fi privacy.*
Retrieved January 22, 2022, from https://support.apple.com/en-ca/guide/security/secb9cb3140c/web

Google. (2021). *Implementing MAC Randomization.*
Retrieved January 22, 2022, from https://source.android.com/devices/tech/connect/wifi-mac-randomization

Wireshark. (2021). *OUI Lookup Tool.*
Retrieved January 22, 2022, from https://www.wireshark.org/tools/oui-lookup.html

Darchis. (2020). *802.11 frames : A starter guide to learn wireless sniffer traces.*
Retrieved January 22, 2022, from https://community.cisco.com/t5/wireless-mobility-documents/802-11-frames-a-starter-guide-to-learn-wireless-sniffer-traces/ta-p/3110019

Alfa. (2021). *AWUS1900 Specifications.*
Retrieved January 22, 2022, from https://www.alfa.com.tw/products/awus1900?variant=36473966231624

Wikipedia. (2021). *List of WLAN channels.*
Retrieved January 22, 2022, from
https://en.wikipedia.org/wiki/List_of_WLAN_channels#2.4_GHz_(802.11b/g/n/ax)

Wireshark. (2021). *Display Filter Reference: IEEE 802.11 Radiotap Capture header*
Retrieved January 22, 2022, from https://www.wireshark.org/docs/dfref/r/radiotap.html

Sonicwall. (2019). *Wireless: SNR, RSSI and Noise basics of wireless troubleshooting.*
Retrieved January 22, 2022, from https://www.sonicwall.com/support/knowledge-base/wireless-snr-rssi-and-noise-basics-of-wireless-troubleshooting/180314090744170/

Eastlake. (2009). *RFC 7402- IANA Considerations and IETF Protocol and Documentation Usage for IEEE 802 Parameters*
Retrieved January 22, 2022, from https://datatracker.ietf.org/doc/html/rfc7042

Lashkari (2016). *Characterization of Encrypted and VPN Traffic Using Time-Related Features*
Retrieved January 22, 2022, from
https://www.researchgate.net/publication/291697471_Characterization_of_Encrypted_and_VPN_Traffic_Using_Time-Related_Features

Williams. (2021). *Pattern of life analysis: how timelines uncover hidden behaviors.*

Retrieved January 22, 2022, from https://cambridge-intelligence.com/pattern-of-life-analysis/#:~:text=Pattern%20of%20life%20analysis%20is,large%20quantities%20of%20observed%20data

6. Appendix A - Event Overlay on Capture

The following figures show the real-world events overlaid on the data capture and signal strength time charts.

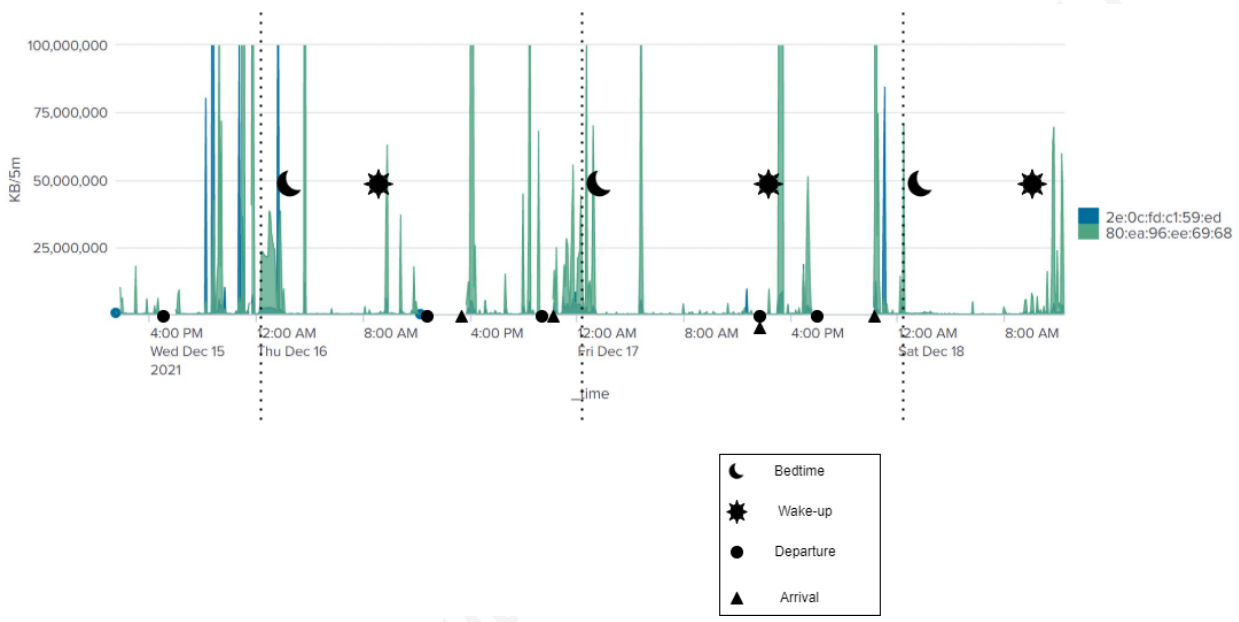


Figure A1: Significant Event Overlay on Data Volume

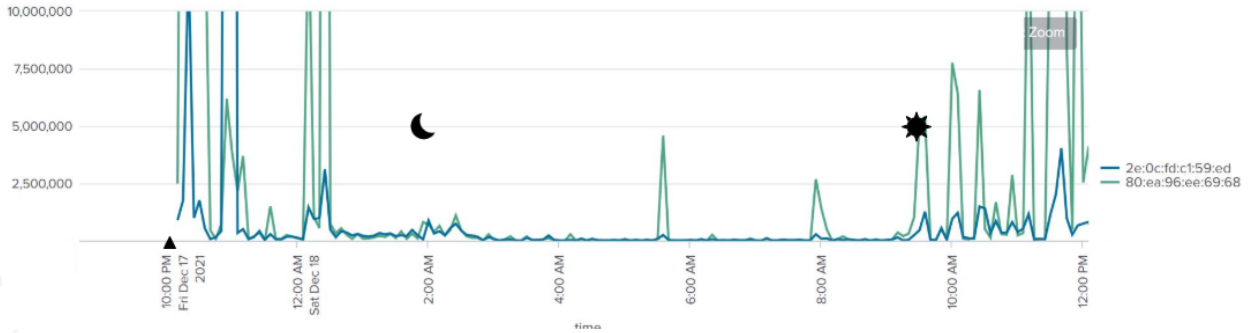
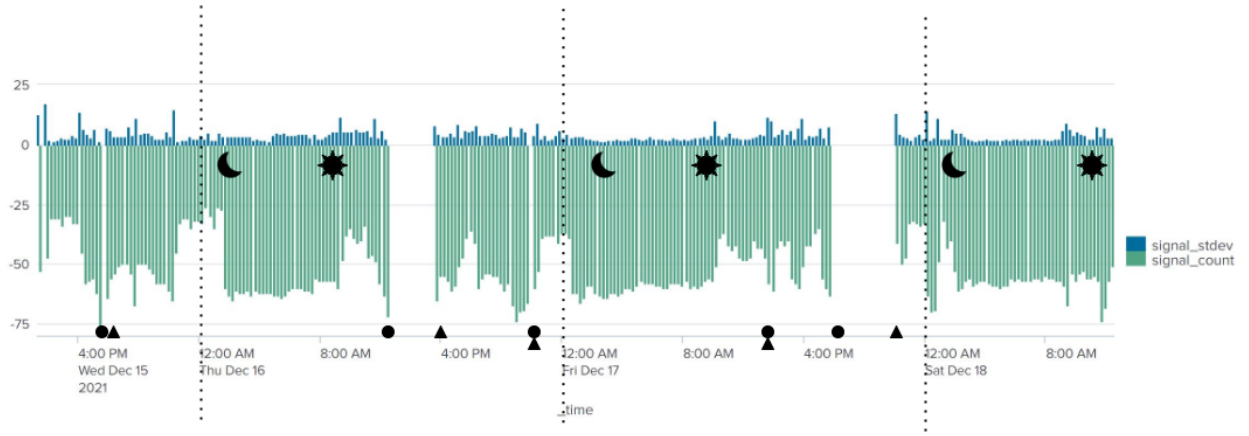


Figure A2: Focused view on Dec 18



Jason Leverton, Leverton.jd@gmail.com

Figure A3: Significant Event Overlay on Signal Strength

December 15

- 1325 : Capture started
- 1703 : Departed area
- 1723 : Home Arrival
- 2255 : iOS update initiates large update

December 16

- 0100 : Bedtime (Time went to bedroom)
- 0900 : Wake up
- 0928 : Office
- 1215 : Departed area
- 1545 : Home Arrival
- 2130 : Departed area
- 2210 : Home arrival

December 17

- 0024 : Bedtime (Time went to bedroom)
- 1013 : Wake up
- 1340 : Departed Area
- 1358 : Home arrival
- 1530 : Departed Area
- 2212 : Home arrival

December 18

- 0201 : Bedtime (Time went to bedroom)

- 0930 : Wake up

High usage Periods:

- December 15, ~20:55
- December 15, ~22:55
- December 16, 1600-1700
- December 17, ~00:30
- December 17, ~22:00
- December 18, ~00:0

7. Appendix B - Timecharts of 15 Minute Baseline Activity

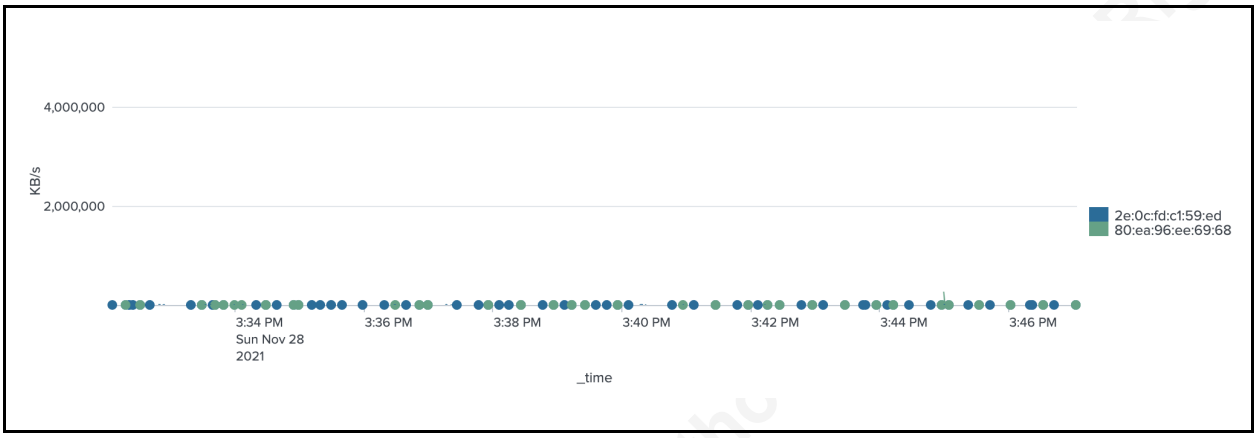


Figure B1: Use for 15 Minutes

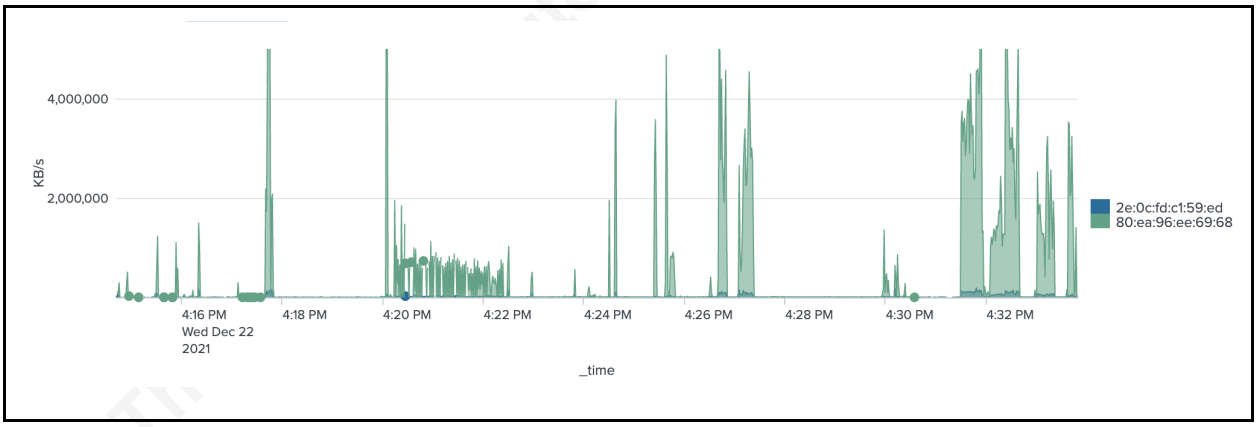
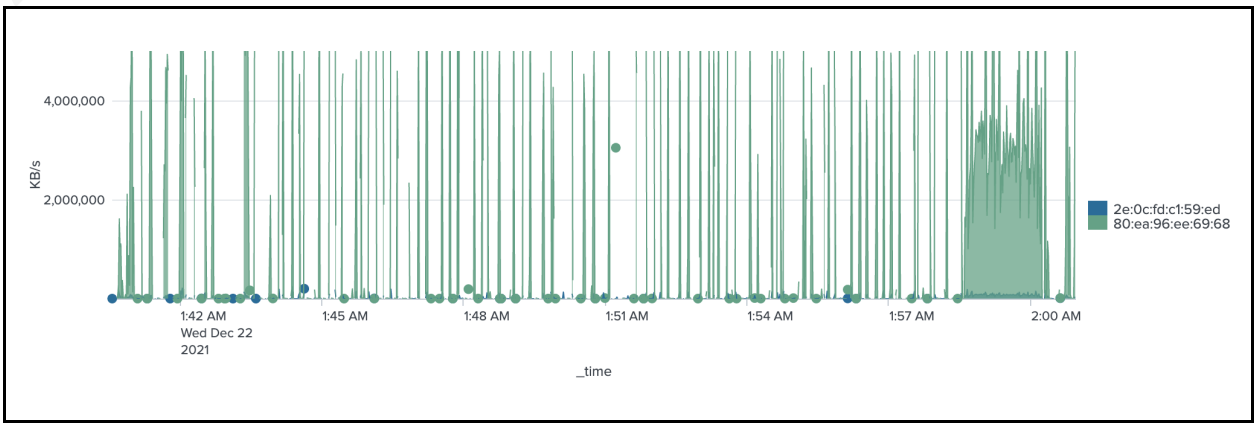


Figure B2: Light Use For 15 Minutes



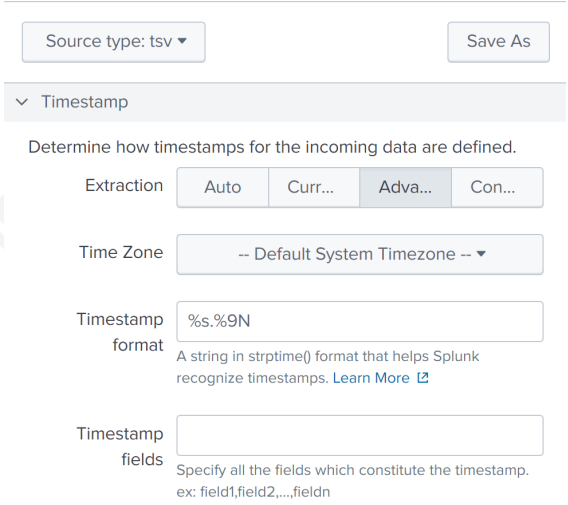
Jason Leverton, Leverton.jd@gmail.com

Figure B3: Streaming Use For 15 Minutes

8. Appendix C - Splunk Ingest Format

The following steps were used to input data into Splunk from the Web App:

1. Data Input
2. Files & Directories
3. New Local File & Directory
4. Index Once (Continuously Monitor if scripted for real-time)
5. SourceType: .tsv
6. Time Extraction: Advanced
    a. %s.%9N (Default epoch time from tshark)



    b.
7. Searching and Reporting
8. Host: Anything
9. Index: Create New Index “Index_a”
    a. Use a new index for this, then searching can just be “Any Time”
10. Review & Submit
11. Search “index=index_a” for “Any Time”.

Jason Leverton, Leverton.jd@gmail.com